Copulas based on Marshall–Olkin machinery

Fabrizio Durante, Stéphane Girard, Gildas Mazo

Copulas based on Marshall–Olkin machinery

Fabrizio Durante* and Stéphane Girard and Gildas Mazo

Abstract We present a general construction principle for copulas that is inspired by the celebrated Marshall–Olkin exponential model. From this general construction method we derive special sub–classes of copulas that could be useful in different situations and recall their main properties. Moreover, we discuss possible estimation strategy for the proposed copulas. The presented results are expected to be useful in the construction of stochastic models for lifetimes (e.g. in reliability theory) or in credit risk models.

1 Introduction

The study of multivariate probability distribution function has been one of the classical topics in the statistical literature once it was recognized at large that the independence assumption cannot describe conveniently the behavior of a random system composed by several components. Since then, different attempts have been done in order to provide more flexible methods to describe the variety of dependence-types that may occur in practice. Unfortunately, the study of high–dimensional models is not that simple when the dimension goes beyond 2 and the range of these mod-

Fabrizio Durante
Faculty of Economics and Management, Free University of Bozen–Bolzano, Bolzano, Italy
e-mail: fabrizio.durante@unibz.it

Stéphane Girard
INRIA, Laboratoire Jean Kuntzmann (LJK), Grenoble, France
e-mail: stephane.girard@inria.fr

Gildas Mazo
INRIA, Laboratoire Jean Kuntzmann (LJK), Grenoble, France
e-mail: gildas.mazo@inria.fr

* The first author acknowledges the support by Faculty of Economics and Management at Free University of Bozen-Bolzano via the project MUD.

els is still not rich enough for the users to choose one that satisfies all the desired properties.

One of the few examples of high–dimensional models that have been used in an ample spectrum of situations is provided by the Marshall–Olkin distribution, introduced in [24] and, hence, developed through various generalizations (as it can be noticed by reading the other contributions to this volume).

The starting point of the present work is to combine the general idea provided by Marshall–Olkin distributions with a copula-based approach. Specifically, we provide a general construction principle, the so–called Marshall–Olkin machinery, that generates many of the families of copulas that have been recently considered in the literature. The methodology is discussed in detail by means of several illustrations. Moreover, possible fitting strategies for the proposed copulas are also presented.

2 Marshall–Olkin machinery

Consider a system composed by $d \geq 2$ components with a random lifetime. We are mainly interested in the deriving an interpretable model for the system supposing that the lifetime of each component may be influenced by adverse factors, commonly indicated as *shocks*. Such shocks can be, for instance, events happening in the environment where the system is working, or simply can be caused by deterioration of one or more components. In a different context, like credit risk, one may think that the system is a portfolio of assets, while the shocks represent arrival times of economic catastrophes influencing the default of one or several assets in the portfolio.

To provide a suitable stochastic model for such situations, let $(\Omega, \mathscr{F}, \mathbb{P})$ be a given probability space.

- For $d \geq 2$, consider the r.v.'s $X_1, \ldots, X_d$ such that each X_i is distributed according to a d.f. F_i, $X_i \sim F_i$. Each X_i can be interpreted as a shock that may effect only the i–th component of the system, i.e. the idiosyncratic shock.
- Let $\mathscr{S} \neq \emptyset$ be a collection of subsets $S \subseteq \{1, 2, \ldots, d\}$ with $|S| \geq 2$, i.e. $\mathscr{S} \subseteq 2^{\{1,2,\ldots,d\}}$. For each $S \in \mathscr{S}$ consider the r.v.'s Z_S with probability d.f. G_S. Such a Z_S can be interpreted as an (external) shock that may affect the stochastic behavior of all the system components with index $i \in S$, i.e. the systemic shock.
- Assume a given dependence among the introduced random vectors $\mathbf{X}$ and $\mathbf{Z}$, i.e. suppose the existence of a given copula C such that, according to Sklar's Theorem [32], one has

$$(\mathbf{X}, \mathbf{Z}) \sim C((F_i)_{i=1,\ldots,d}, (G_S)_{S \in \mathscr{S}}). \tag{1}$$

  The copula C describes how the shocks $\mathbf{X}$ and $\mathbf{Z}$ are related each other.
- For $i = 1, \ldots, d$, assume the existence of a linking function ψ_i that expresses how the effects produced by the shock X_i and all the shocks Z_S with $i \in S$ are combined together and acts on the i–th component.

Given all these assumptions, we call *Marshall–Olkin machinery* any d–dimensional stochastic model $\mathbf{Y} = (Y_1, \ldots, Y_d)$ that has the stochastic representation:

$$Y_i = \psi_i(X_i, Z_{S:\, i \in S}). \tag{2}$$

Such a general framework includes most of the so–called shock models presented in the literature. Notably, Marshall–Olkin multivariate (exponential) distribution is simply derived from the previous framework by assuming that

$$(\mathbf{X}, \mathbf{Z}) \sim \left(\prod_{i=1}^{d} F_i \right) \cdot \left(\prod_{\emptyset \neq S \in 2^{\{1,2,\ldots,d\}}} G_S \right), \tag{3}$$

i.e. all the involved r.v.'s are mutually independent, each X_i and each Z_S have exponential survival distribution, $\psi_i = \max$.

However, it includes also various Marshall–Olkin type generalized families, including, for instance, the family presented in [19] that is obtained by assuming that X_i's are not identically distributed (see also [28]).

By suitable modifications, Marshall–Olkin machinery can be adapted in order to obtain general construction methods for copulas. In fact, the growing use of copulas in applied problems always requires the introduction of novel families that may underline special features like tail dependence, asymmetries, etc. Specifically, in order to ensure that the distribution function of $(Y_1, \ldots, Y_d)$ of Eq. (2) is a copula it could be convenient to select all X_i's and all G_S's with support on $[0,1]$ and, in addition, ψ_i with range in $[0,1]$. Obviously, one has also to check that each Y_i is uniformly distributed in $[0,1]$. We call *Marshall–Olkin machinery* any construction methods for copulas that is based on previous arguments. In the following we are interested in presenting some specific classes generated by this mechanism.

Provided that the copula C and the marginal d.f.'s of Eq. (1) can be easily simulated, distribution functions (in particular, copulas) generated by Marshall–Olkin machinery can be easily simulated. However, if no constraints are require on the choice of $\mathscr{S}$, such distributions are specified by (at least) 2^d parameters, namely

- d parameters related to X_i's;
- $2^d - d - 1$ parameters related to Z_S's;
- (at least) one parameter related to the copula C.

Hence, such a kind of model soon becomes unhandy as the dimension increases. Therefore, we are interested in flexible subclasses generated by Marshall–Olkin machinery with fewer parameters that are better suited for high-dimensional applications.

3 Copulas generated by one independent shock

To provide a preliminary class generated by Marshall–Olkin mechanism, consider the case when the system is subjected to individual shocks and one global shock that is independent of the previous ones. In such a case, copulas may be easily obtained in view of the following result.

Theorem 1. *For $d \geq 2$, consider the continuous r.v. $\mathbf{X} = (X_1, \dots, X_d)$ having copula C and such that each X_i is distributed according to a d.f. F supported on $[0,1]$. Consider the r.v. Z with probability d.f. G such that Z is independent of $\mathbf{X}$. For every $i = 1, \dots, d$, set*

$$Y_i := \max\{X_i, Z\}.$$

If $G(t) = t/F(t)$ for $t \in]0,1]$, then the d.f. of $(Y_1, \dots, Y_d)$ is a copula, given by

$$\widetilde{C}(\mathbf{u}) = G(u_{(1)}) \cdot C(F(u_1), \dots, F(u_d)), \tag{4}$$

where $u_{(1)} = \min_i u_i$.

Proof. The expression of $\widetilde{C}$ can be obtained by direct calculation. Moreover, since $\widetilde{C}$ is obviously a d.f., the proof consists of showing that the univariate margins of $\widetilde{C}$ are uniform on $(0,1)$. However, this is a straightforward consequence of the equality $F(t)G(t) = t$ on $(0,1)$. $\square$

Models of type (4) can be also deduced from [29].

Remark 1. In the assumption of Theorem 1, since G has to be a d.f. it follows that $t \leq F(t)$ for all $t \in [0,1]$. Moreover, the condition $t/F(t)$ being increasing is equivalent (assuming differentiability of F) to $(\log(t))' \geq (\log(F(t)))'$ on $(0,1)$. Finally, notice that if F is concave, then $t \mapsto t/F(t)$ is increasing on $(0,1)$ (see, e.g., [25]).

Remark 2. It is worth noticing that the copula $\widetilde{C}$ in (4) can be rewritten as

$$\widetilde{C}(\mathbf{u}) = \min(G(u_1), \dots, G(u_d)) \cdot C(F(u_1), \dots, F(u_d)).$$

Intuitively, it is the product of the comonotonicity copula $M_d(\mathbf{u}) = \min\{u_1, u_2, \dots, u_d\}$ and the copula C with some suitable transformation of the respective arguments. This way of combining copulas was considered, for the bivariate case, in [4, 12], and for the general case in [20, Theorem 2.1].

As it can be seen from Figure 1, the main feature of copulas of type (4) is that they have a singular component along the main diagonal of the copula domain $[0,1]^d$. In general, if the r.v. $\mathbf{X}$ has copula $\widetilde{C}$ of type (4), then $\mathbb{P}(X_1 = X_2 = \cdots = X_d) > 0$. This feature could be of great interest when the major issue is to model a vector of lifetimes and it is desirable that defaults of two or more components may occur at the same time with a non-zero probability.

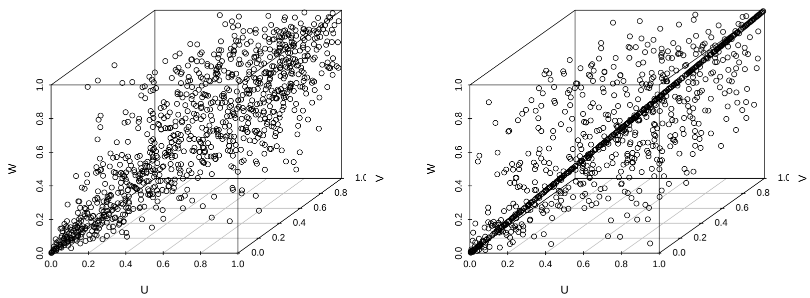

Fig. 1 Trivariate Clayton copula (left) and its modification of type (4) with $F(t) = t^{1-\alpha}$, $\alpha = 0.60$ (right).

Roughly speaking, a model of type (4) tends to increase the positive dependence. In fact, since F is a d.f. such that $F(t) \geq t$ on $[0,1]$, $\widetilde{C} \geq C$ pointwise, which corresponds to the positive lower orthant dependent order between copulas (see, e.g., [16]). However, $\widetilde{C}$ need not be positive dependent, i.e. $C \geq \Pi_d$ pointwise. For instance, consider the random sample from the copula described in Figure 2. As can be noticed, there is no mass probability around the point $(0,0)$ and, hence, such a copula cannot be greater than Π_2.

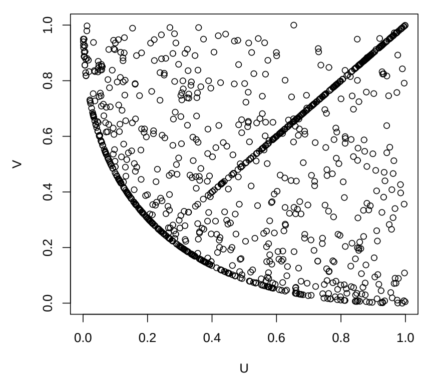

Fig. 2 Copula of type (4) with one shock generated by $F(t) = t^{1-\alpha}$, $\alpha = 0.50$, and C equal to Fréchet lower bound copula W_2.

3.1 The bivariate case

Now, consider the simple bivariate case related to copulas of Theorem 1 by assuming, in addition, that C equals the independence copula Π_2. Specifically, we assume that there exist three independent r.v.'s X_1, X_2, Z whose support is contained in $[0,1]$ such that $X_i \sim F$, $i = 1,2$, and $Z \sim G(t) = t/F(t)$. For $i = 1,2$, we define the new stochastic model

$$Y_i = \max(X_i, Z).$$

Then the d.f. of $\mathbf{Y}$ is given by

$$\widetilde{C}(u_1,u_2) = \min(u_1,u_2)F(\max(u_1,u_2)), \tag{5}$$

Copulas of this type may be rewritten in the form

$$\widetilde{C}(u_1,u_2) = \min(u_1,u_2)\frac{\delta(\max(u_1,u_2))}{\max(u_1,u_2)} \tag{6}$$

where $\delta(t) = \widetilde{C}(t,t)$ is the so-called *diagonal section of* $\widetilde{C}$ (see, for instance, [6]). We refer to [1] for other re-writings. As known, if (U_1,U_2) are distributed according to a copula C, then the diagonal section of C contains the information about the order statistics $\min(U_1,U_2)$ and $\max(U_1,U_2)$. In fact, for every $t \in [0,1]$

$$\mathbb{P}(\max(U_1,U_2) \le t) = \delta(t),$$
$$\mathbb{P}(\min(U_1,U_2) \le t) = 2t - \delta(t).$$

Since the d.f. F related to our shock model equals $\delta(t)/t$ on $(0,1]$, it follows that it determines the behaviour of order statistics of (U_1,U_2). In the case of lifetimes models, this is equivalent to say that the survival of a (bivariate) system is completely driven by one single shock F.

The equivalence of the formulations (5) and (6) suggests two possible ways for constructing a bivariate models of type (4) by either assigning F or δ. In both cases, additional assumptions must be given in order to ensure that the obtained model describes a bona fide copula. These conditions are illustrated here (for the proof, see [8]).

Theorem 2. *Let $\widetilde{C}$ be a function of type* (4)*. Set $\delta := F(t)/t$ on $(0,1]$. Then $\widetilde{C}$ is a copula if, and only if, the functions φ_δ, $\eta_\delta : (0,1] \to [0,1]$ given by*

$$\varphi_\delta(t) := \frac{\delta(t)}{t}, \qquad \eta_\delta(t) := \frac{\delta(t)}{t^2}$$

are increasing and decreasing, respectively.

Notice that both the independence copula $\Pi_2(u_1,u_2) = u_1u_2$ and the comonotonicity copula $M_2(u_1,u_2) = \min(u_1,u_2)$ are examples of copulas of type (5), generated by $F(t)=t$ and $F(t)=1$, respectively. Moreover, an algorithm for simulating such copulas is illustrated in [11, Algorithm 1]. Related random samples are depicted in Figure 3. Another example of copulas of type (5) is given by the bivariate Sato copula of [21], generated by $F(t)=(2-t^{1/\alpha})^{-\alpha}$ for every $\alpha>0$.

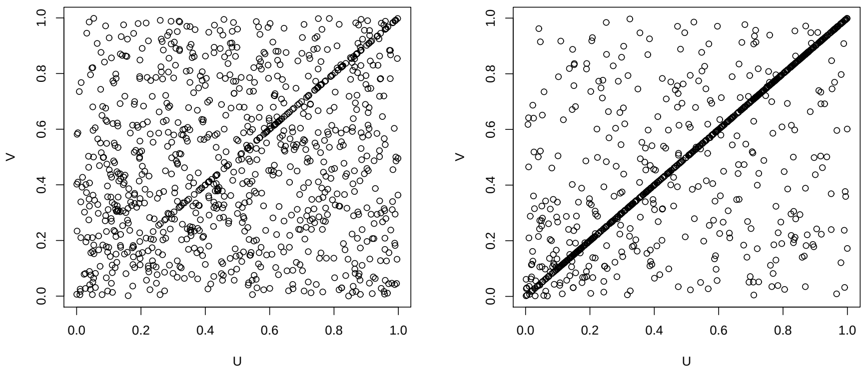

Fig. 3 Copulas of type (4) generated by $F(t)=t^{1-\alpha}$ with $\alpha=0.25$ (left) and $\alpha=0.75$ (right). These are members of Cuadras-Augé family of copulas.

Copulas of type (5) can be interpreted as the exchangeable (i.e., invariant under permutation of their arguments) members of the family proposed in [23, Proposition 3.1]. Since this latter reference was the first work that has explicitly provided sufficient conditions to obtain copulas of type (5), they can also be referred to as *exchangeable Marshall copulas* (shortly, EM copulas), as done in [9]. Notice that EM copulas are also known under the name *semilinear copulas*, a term used in [8], and justified by the fact that the these copulas are linear along suitable segments of their domains (see Figure 4).

EM copulas can model positive quadrant dependence, i.e. each EM copula is greater than Π_2 pointwise. Actually, they even satisfy the stronger positive dependence notion called TP2 (see [5]). Following [7] this implies that, if (X,Y) is an exchangeable vector with EM copula, then the vector of residual lifetimes $(X,Y \mid X>t, Y>t)$ at time $t>0$ is also TP2 and, a fortiori, positive quadrant dependent. Roughly speaking, the positive dependence between the residual lifetimes of the system is (qualitatively) preserved at the increase of age.

Another feature of interest in EM copulas is a kind of stability of this class with respect of certain operation. Usually, risk estimation procedures require the calculation of risk functions (like Value-at-Risk) with respect to some specific information about the dependence. In particular, in this respect, upper and lower bounds for copulas with some specified feature are relevant (see, e.g., [17, 2, 33]). Now, the class of all EM copulas is both a convex and log-convex set in the class of all bivariate

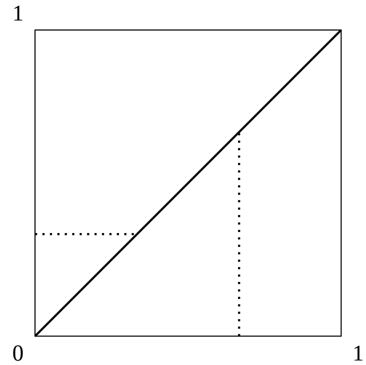

Fig. 4 The dotted lines indicates the typical segments where the restriction of the EM copula to these sets is linear.

copulas. Moreover, it is also closed under pointwise suprema and infima operations. Just to provide an example, notice that the class of bivariate Archimedean copulas is neither convex nor closed under suprema and infima.

3.2 The multivariate case

Copulas of type (5) can be easily extended in any dimension. In fact, consider the general case related to copulas of Theorem 1 by assuming, in addition, that C equals the independence copula Π_2. Specifically, we assume that there exist $(d+1)$ independent r.v.'s $X_1, X_2, \dots, X_d, Z$ whose support is contained in $[0,1]$ such that $X_i \sim F$, $i = 1,2$, and $Z \sim G(t) = t/F(t)$. For $i = 1,2,\dots,d$, we define the new stochastic model $\mathbf{Y}$, where $Y_i = \max(X_i, Z)$. Then the d.f. of $\mathbf{Y}$ is given by

$$\widetilde{C}(\mathbf{u}) = u_{[1]} \prod_{i=2}^{d} F(u_{[i]}), \tag{7}$$

where $u_{[1]}, \dots, u_{[d]}$ denote the components of $(u_1, \dots, u_d)$ rearranged in increasing order. Since $\mathbf{Y}$ has uniform univariate marginals, $\widetilde{C}$ is a copula. Moreover, the following characterization holds (see [10]).

Theorem 3. *Let $F\colon [0,1] \to [0,1]$ be a continuous d.f., and, for every $d \geq 2$, let $\widetilde{C}$ be the function defined by* (7)*. Then $\widetilde{C}$ is a d–copula if, and only if, the function $t \to \frac{F(t)}{t}$ is decreasing on $(0,1]$.*

Example 1. Let $\alpha \in [0,1]$ and consider $F_\alpha(t) = \alpha t + \overline{\alpha}$, with $\overline{\alpha} := 1 - \alpha$. Then $\widetilde{C}_{F_\alpha}$ of type (7) is given by

$$\widetilde{C}_{F_\alpha}(\mathbf{u}) = u_{[1]} \prod_{i=2}^{d} (\alpha u_{[i]} + \overline{\alpha}).$$

In particular, for $d = 2$, we obtain a convex combination of the copulas Π_2 and M_2.

Example 2. Let $\alpha \in [0,1]$ and consider the function $F_\alpha(t) = t^\alpha$. Then $\widetilde{C}_{F_\alpha}$ of type (7) is given by

$$\widetilde{C}_{F_\alpha}(\mathbf{u}) = (\min(u_1, u_2, \dots, u_n))^{1-\alpha} \prod_{i=1}^{d} u_i^\alpha.$$

It generalizes the *Cuadras-Augé family* of bivariate copulas [3]. Further generalization of this family is also included in [22].

Copulas of type (7) have some distinguished features. First, they are exchangeable, a fact that could represented a limitation in some applications. Second, their tail behavior is only driven by the generator function F. To make this statement precise, consider the following extremal dependence coefficient introduced in [13].

Definition 1. Let $\mathbf{X}$ be a random vector with univariate margins $F_1, \dots, F_d$. Let $F_{\min} := \min_i F_i(X_i)$ and $F_{\max} := \max_i F_i(X_i)$. The *lower extremal dependence coefficient* (LEDC) and the *upper extremal dependence coefficient* (UEDC) of $\mathbf{X}$ are given, respectively, by

$$\varepsilon_L := \lim_{t \to 0^+} P[F_{\max} \le t | F_{\min} \le t], \quad \varepsilon_U := \lim_{t \to 1^-} P[F_{\min} > t | F_{\max} > t],$$

if the limits exist.

Notice that, in the bivariate case, LEDC and UEDC closely related to the lower and upper tail dependence coefficients (write: LTDC and UTDC, respectively), which are given by

$$\lambda_L = \lim_{t \to 0^+} \frac{C(t,t)}{t} \quad \text{and} \quad \lambda_U = \lim_{t \to 1^-} \frac{1 - 2t + C(t,t)}{1-t}.$$

The following result holds [10]. Notice that non–trivial LEDC occurs only when F is discontinuous at 0.

Theorem 4. *Let $\widetilde{C}$ be a copula of type* (7) *generated by a differentiable F. Then, the LEDC and UEDC are, respectively, given by*

$$\varepsilon_L = \frac{(F(0^+))^{n-1}}{\sum_{i=1}^{n} (-1)^{i-1} \binom{n}{i} (F(0^+))^{i-1}}, \qquad \varepsilon_U = \frac{1 - F'(1^-)}{1 + (n-1) F'(1^-)}.$$

Although copulas of type (7) seem a quite natural generalization of EM copulas, for practical purposes their main inconvenience is that only one function F describes the d–dimensional dependence.

To overcome such oversimplification, a convenient generalization has been provided in [21]. Basically, copulas of type (7) have been extended to the form

$$\widetilde{C}(\mathbf{u}) = u_{[1]} \prod_{i=2}^{d} F_i(u_{[i]}), \tag{8}$$

for suitable functions $F_2, \dots, F_d$. Interestingly, a subclass of the considered copulas also can be interpreted in terms of exceedance times of an increasing additive stochastic process across independent exponential trigger variables.

In order to go beyond exchangeable models, keeping a certain tractability and/or simplicity of the involved formulas, a possible strategy could be to combine in a suitable way pairwise copulas of type (7) in order to build up a general model. This is pursued, for instance, in [11] and [26], by using as building block Marshall–Olkin copulas. Both procedures are described in Section 4.

4 Combining Marshall-Olkin bivariate copulas to get flexible multivariate models

Motivated by the fact that bivariate dependencies are not difficult to check out, it may be of interest to construct a multivariate copula such that each of its bivariate margins depends upon a suitable parameter. For example, following [11], one could introduce a multivariate (extreme-value) copula such that each bivariate marginal C_{ij} belongs to the Cuadras-Augé family:

$$C_{ij}(u_i, u_j) = \Pi_2(u_i, u_j)^{1-\lambda_{ij}} M_2(u_i, u_j)^{\lambda_{ij}}.$$

To this end, following a Marshall–Olkin machinery, one may consider the following stochastic representation of r.v.'s whose support is contained in $[0,1]$:

- For $d \geq 2$, consider the r.v.'s $X_1, \dots, X_d$ such that each X_i is distributed according to a d.f. $F_i(t) = t^{1-\sum_{j\neq i}\lambda_{ij}}$ for $i = 1, 2, \dots, d$.
- For $i, j \in \{1, 2, \dots, d\}$, $i < j$, consider the r.v. Z_{ij} distributed according to $G_{ij}(t) = t^{\lambda_{ij}}$.
- We assume independence among all X's and all Z's.

For $i = 1, 2, \dots, d$, we define the new r.v. $\mathbf{Y}$ whose components are given by

$$Y_i = \max\left(X_i, Z_{i1}, \dots, Z_{i(j-1)}, Z_{i(j+1)}, \dots, Z_{id}\right).$$

Basically, Y_i is determined by the interplay among the individual shock X_i and all the pairwise shocks related to the i–th component of a system. Then the d.f. of $\mathbf{Y}$ is given by

$$C^{\mathbf{pw}}(\mathbf{u}) = \prod_{i=1}^{d} u_i^{1-\sum_{j\neq i}\lambda_{ij}} \prod_{i<j} (\min\{u_i, u_j\})^{\lambda_{ij}}.$$

Here, for every $i, j \in \{1, \dots, d\}$ and $i < j$, $\lambda_{ij} \in [0,1]$ and $\lambda_{ij} = \lambda_{ji}$. Moreover, if, for every $i \in \{1, \dots, d\}$, $\sum_{j=1, j\neq i}^{d} \lambda_{ij} \leq 1$, then $C^{\mathbf{pw}}$ is a multivariate d-copula (that is also an extreme–value copula). This copula has Cuadras-Augé bivariate margins and, therefore, may admit non–zero UTDCs. However, even if this model is nonexchangeable, the constraints given on the parameters are a severe drawback. For instance, when $d = 3$, one of the constraint is that $\lambda_{12} + \lambda_{13} \leq 1$. Therefore, the two pairs (X_1, X_2) and (X_1, X_3) cannot have a large UTDC together. In the simplified case where all the parameters λ_{ij}'s are equal to a common value $\lambda \in [0,1]$, the copula $C^{\mathbf{pw}}$ reduces to

$$C^{\mathbf{pw}}(\mathbf{u}) = \prod_{i=1}^{d} u_{[i]}^{1-\lambda(i-1)}$$

with the constraint $\lambda \leq \frac{1}{d-1}$.

Another way of combining Marshall-Olkin bivariate copulas, that does not suffer from any constraints, and that still yield a flexible model, was hence proposed in [26] and it is given next.

Let $Y_0, Y_1, \ldots, Y_d$ be standard uniform random variables such that the coordinates of $(Y_1, \ldots, Y_d)$ are conditionally independent given Y_0. The variable Y_0 plays the role of a latent, or unobserved, factor. Let us write C_{0i} the distribution of (Y_0, Y_i) and $C_{i|0}(\cdot|u_0)$ the conditional distribution of Y_i given $Y_0 = u_0$, for $i = 1, \ldots, d$. The copulas C_{0i} are called the *linking copulas* because they link the factor Y_0 to the variables of interest Y_i. It is easy to see that the distribution of $(Y_1, \ldots, Y_d)$ is given by the so called *one-factor copula* [18]

$$C(\mathbf{u}) = \int_0^1 C_{1|0}(u_1|u_0) \ldots C_{d|0}(u_d|u_0)\, du_0. \tag{9}$$

When one chooses C_{0i} to be of type (5) with generator F_i, the integral (9) can be calculated. This permits to exhibit interesting properties for this class of copulas. Thus, calculating the integral yields

$$\begin{aligned} C(\mathbf{u}) = u_{(1)} \Bigg[& \left(\prod_{j=2}^{d} u_{(j)}\right) \int_{u_{(d)}}^{1} \prod_{j=1}^{d} F_j'(x)dx + F_{(1)}(u_{(2)}) \left(\prod_{j=2}^{d} F_{(j)}(u_{(j)})\right) \\ & + \sum_{k=3}^{d} \left(\prod_{j=2}^{k-1} u_{(j)}\right) \left(\prod_{j=k}^{d} F_{(j)}(u_{(j)})\right) \int_{u_{(k-1)}}^{u_{(k)}} \prod_{j=1}^{k-1} F_{(j)}'(x)dx \Bigg], \end{aligned} \tag{10}$$

where $F_{(i)} := F_{\sigma(i)}$ and σ is the permutation of $(1, \ldots, d)$ such that $u_{\sigma(i)} = u_{(i)}$. The particularity of this copula lies in the fact that it depends on the generators through their reordering underlain by the permutation σ. This feature gives its flexibility to the model. Observe also that $C(\mathbf{u})$ writes as $u_{(1)}$ multiplied by a functional of $u_{(2)}, \ldots, u_{(d)}$, form that is similar to (7). Interestingly, all the bivariate copulas derived from this model have a simple form as stated below.

Proposition 1. *Let C_{ij} be a bivariate margin of* (10)*. Then C_{ij} is a copula of type* (5) *with generator*

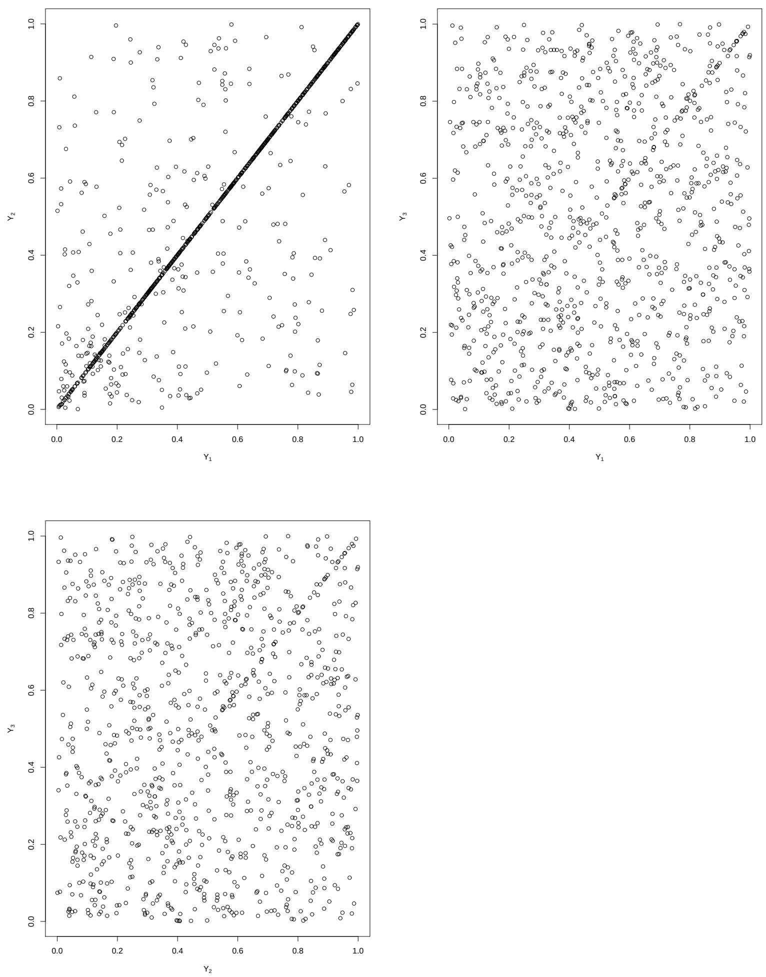

Fig. 5 Random sample of 1000 points from a 3–copula of type (10) with Cuadras-Augé generators with parameters $(\alpha_1, \alpha_2, \alpha_3) = (0.9, 0.9, 0.1)$. The figure shows the three bivariate margins.

$$F_{ij}(t) = F_i(t)F_j(t) + t\int_t^1 F_i'(x)F_j'(x)dx.$$

By Proposition 1, the class of copulas (10) can be viewed as a generalization in higher dimension of the bivariate copulas of type (5). Moreover, the LTDC and UTDC coefficients are given by

$$\lambda_{L,ij} = F_i(0)F_j(0) \text{ and } \lambda_{U,ij} = (1 - F_i'(1^-))(1 - F_j'(1^-)).$$

Example 3 (Fréchet generators). Let $F_i(t) = \alpha_i t + 1 - \alpha_i$, $\alpha_i \in [0,1]$. By Proposition 1, F_{ij} is given by

$$F_{ij}(t) = (1-(1-\alpha_i)(1-\alpha_j))\,t + (1-\alpha_i)(1-\alpha_j).$$

The LTDC and UTDC are respectively given by

$$\lambda_{L,ij} = \lambda_{U,ij} = (1-\alpha_i)(1-\alpha_j).$$

Example 4 (Cuadras-Augé generators). Let $F_i(t) = t^{\alpha_i}$, $\alpha_i \in [0,1]$. By Proposition 1, F_{ij} is given by

$$F_{ij}(t) = \begin{cases} t^{\alpha_i+\alpha_j}\left(1-\frac{\alpha_i\alpha_j}{\alpha_i+\alpha_j-1}\right)+t\,\frac{\alpha_i\alpha_j}{\alpha_i+\alpha_j-1} & \text{if } \alpha_i+\alpha_j \neq 1 \\ t(1-(1-\alpha)\alpha\log t) & \text{if } \alpha = \alpha_j = 1-\alpha_i. \end{cases}$$

The LTDC and UTDC are respectively given by

$$\lambda_{L,ij} = 0 \text{ and } \lambda_{U,ij} = (1-\alpha_i)(1-\alpha_j).$$

In the case $d = 3$, Figure 5 depicts a simulated sample of 1000 observations from this copula with parameter $(\alpha_1, \alpha_2, \alpha_3) = (0.9, 0.9, 0.1)$.

Unlike copulas of type (7), the copulas of type (10) are not exchangeable. They are determined by d generators $F_1, \ldots, F_d$, which combine together to give a more flexible dependence structure. Taking various parametric families, as illustrated in Example 3 and 4, allows to obtain various tail dependencies.

5 Some comments about statistical inference procedures

The construction principle presented above provides copulas that are not absolutely continuous (up to trivial cases) with respect to the restriction of the Lebesgue measure to the copula domain. Thus, statistical procedures that requires density of the related distribution can not be applied. Moreover, the presence of a singular component often causes the presence of points where the derivatives do not exist, a fact that should also considered for the direct applicability of statistical techniques based on moments' method (see, for instance, [14]).

In this section, we present instead a method to estimate the parameters of the copulas encountered in this paper, which is based on some recent results in [27].

Let

$$(X_1^{(1)}, \ldots, X_d^{(1)}), \ldots, (X_1^{(n)}, \ldots, X_d^{(n)})$$

be a sample of n independent and identically distributed d-variate observations from $(X_1, \ldots, X_d)$, a random vector distributed according to F and with copula C, where $C \equiv C_\theta$ belongs to a parametric family indexed by a parameter vector $\theta \in \Theta \in \mathbb{R}^q$, $q \leq d$. The estimator is defined as

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \left(\hat{r} - r(\theta)\right)^T \widehat{W} \left(\hat{r} - r(\theta)\right), \tag{11}$$

where $\hat{r} = (\hat{r}_{1,2}, \dots, \hat{r}_{d-1,d})$, $r(\theta) = (r_{1,2}(\theta), \dots, r_{d-1,d}(\theta))$ and $\widehat{W}$ is a positive definite (weight) matrix with full rank; the coordinate $r_{i,j}(\theta)$ is to be replaced by a dependence coefficient between X_i and X_j, and $\hat{r}_{i,j}$ by its empirical estimator – for instance, the Spearman's rho or the Kendall's tau. The approach (11) can be viewed as an extension to the multivariate case of the Spearman's rho / Kendall's tau inversion method [15]. The asymptotic properties of $\hat{\theta}$ have been studied in [27] in the case where the copulas do not have partial derivatives, as it is the case of the copulas in this article. In particular, it was shown that, under natural identifiability conditions on the copulas, $\hat{\theta}$ exists, is unique with probability tending to 1 as $n \to \infty$, and in that case, is consistent and $\sqrt{n}(\hat{\theta} - \theta)$ tends to a Gaussian distribution.

For the purpose of illustration, we present here a real-data application of the method by using a dataset consisting of 3 gauge stations where annual maximum flood data were recorded in northwestern Apennines and Thyrrhenian Liguria basins (Italy): Airole, Merelli, and Poggi. The dataset is the same used in [11] to which we refer for more detailed description.

In order to fit the dependence among these three gauge stations, we use the class of copulas given by

$$C(u_1, u_2, u_3) = \left(\prod_{i=1}^{3} u_i^{1-\theta_i}\right) \min_{i=1,2,3} (u_i^{\theta_i}), \quad \theta_i \in [0,1], i = 1,2,3.$$

Such a copula can be also seen as generated by Marshall–Olkin machinery, by assuming that $\mathbf{X}$ and $\mathbf{Z}$ are independent r.v.'s of length d whose copula is given by Π_d and M_d, respectively, F_i and G_i are power functions, and $\psi = \max$.

The estimator (11) coordinates $\hat{\theta}_1, \hat{\theta}_2$ and $\hat{\theta}_3$ are given by

$$\hat{\theta}_i = \frac{1}{2} \left(1 + \frac{1}{\hat{\tau}_{i,j}} + \frac{1}{\hat{\tau}_{i,k}} - \frac{1}{\hat{\tau}_{j,k}}\right),$$

where the $\hat{\tau}_{i,j}$ are the pairwise sample Kendall's τ coefficients. Notice that, as consequence of [27, Proposition 2], when the number of parameters is equal to the number of pairs $d(d-1)/2$, then the estimator given by (11) does not depend on the weights.

These previous estimates help to quantify the critical levels and return periods corresponding to this dataset (see, e.g., [30, 31]). In hydrology, a critical level p corresponding to a return period T is defined through the relationship

$$T = \frac{1}{1 - \mathbb{P}(C(F_1(Y_1), \dots, F_d(Y_d)) \leq p)}, \quad p \in [0,1],$$

where $Y_1, \dots, Y_d$ are the r.v.'s of interest and $F_1, \dots, F_d$ their respective univariate marginals. The return period can be interpreted as the average time elapsing between two dangerous events. For instance, $T = 30$ years means that the event happens once every 30 years in average. Figure 6 shows the estimated critical levels, along with confidence intervals, associated to the fitted dataset.

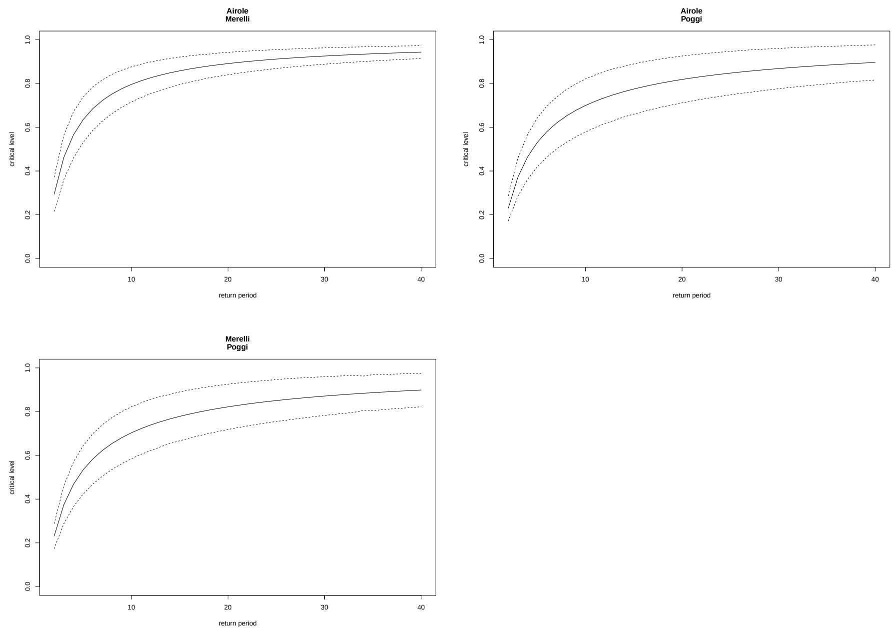

Fig. 6 Critical levels for $T = 2, \ldots, 40$ together with 95% confidence intervals.

6 Conclusions

We have presented a construction principle of copulas that is inspired by the seminal Marshall–Olkin idea of constructing shock models. The copulas obtained in this way have some distinguished properties:

- they have an interpretation in terms of (local or global) shocks;
- they enlarge known families of copulas by including asymmetric copula (in the tails) and/or non–exchangeability;
- they have a natural sampling strategy;
- they can be used to build models with singular components, a fact that is useful when modeling joint defaults of different lifetimes (i.e. credit risk).
- they can be fitted to real data with simple novel methodology.

We think that all these properties make these constructions appealing in several applications.

References

1. Amblard, C., Girard, S.: A new extension of bivariate FGM copulas. Metrika **70**, 1–17 (2009)
2. Bernard, C., Liu, Y., MacGillivray, N., Zhang, J.: Bounds on capital requirements for bivariate risk with given marginals and partial information on the dependence. Dependence Modeling **1**, 37–53 (2013)
3. Cuadras, C.M., Augé, J.: A continuous general multivariate distribution and its properties. Comm. Statist. A—Theory Methods **10**(4), 339–353 (1981)
4. Durante, F.: Generalized composition of binary aggregation operators. Internat. J. Uncertain. Fuzziness Knowledge-Based Systems **13**(6), 567–577 (2005)
5. Durante, F.: A new class of symmetric bivariate copulas. J. Nonparametr. Stat. **18**(7-8), 499–510, (2007) (2006)
6. Durante, F., Fernández-Sánchez, J., Pappadà, R.: Copulas, diagonals and tail dependence. Fuzzy Sets and Systems (2015). In press
7. Durante, F., Foschi, R., Spizzichino, F.: Threshold copulas and positive dependence. Statist. Probab. Lett. **78**(17), 2902–2909 (2008)
8. Durante, F., Kolesárová, A., Mesiar, R., Sempi, C.: Semilinear copulas. Fuzzy Sets and Systems **159**(1), 63–76 (2008)
9. Durante, F., Okhrin, O.: Estimation procedures for exchangeable Marshall copulas with hydrological application. Stoch. Environ. Res Risk Asess **29**(1), 205–226 (2015)
10. Durante, F., Quesada-Molina, J.J., Úbeda-Flores, M.: On a family of multivariate copulas for aggregation processes. Inform. Sci. **177**(24), 5715–5724 (2007)
11. Durante, F., Salvadori, G.: On the construction of multivariate extreme value models via copulas. Environmetrics **21**(2), 143–161 (2010)
12. Durante, F., Sempi, C.: On the characterization of a class of binary operations on bivariate distribution functions. Publ. Math. Debrecen **69**(1-2), 47–63 (2006)
13. Frahm, G.: On the extremal dependence coefficient of multivariate distributions. Statist. Probab. Lett. **76**(14), 1470–1481 (2006)
14. Genest, C., Carabarím-Aguirre, A., Harvey, F.: Copula parameter estimation using Blomqvist's beta. J. SFdS **154**(1), 5–24 (2013)
15. Genest, C., Favre, A.C.: Everything you always wanted to know about copula modeling but were afraid to ask. J. Hydrol. Eng. **12**(4), 347–368 (2007)
16. Joe, H.: Dependence Modeling with Copulas. Chapman & Hall/CRC, London (2014)
17. Kaas, R., Laeven, R.J.A., Nelsen, R.B.: Worst VaR scenarios with given marginals and measures of association. Insurance Math. Econom. **44**(2), 146–158 (2009)
18. Krupskii, P., Joe, H.: Factor copula models for multivariate data. J. Multivariate Anal. **120**, 85–101 (2013)
19. Li, X., Pellerey, F.: Generalized Marshall–Olkin distributions and related bivariate aging properties. J. Multivariate Anal. **102**(10), 1399–1409 (2011)
20. Liebscher, E.: Construction of asymmetric multivariate copulas. J. Multivariate Anal. **99**(10), 2234–2250 (2008)
21. Mai, J.F., Schenk, S., Scherer, M.: Exchangeable exogenous shock models. Bernoulli (2015). In press
22. Mai, J.F., Scherer, M.: Lévy-Frailty copulas. J. Multivariate Anal. **100**(7), 1567–1585 (2009)
23. Marshall, A.W.: Copulas, marginals, and joint distributions. In: Distributions with fixed marginals and related topics (Seattle, WA, 1993), *IMS Lecture Notes Monogr. Ser.*, vol. 28, pp. 213–222. Inst. Math. Statist., Hayward, CA (1996)
24. Marshall, A.W., Olkin, I.: A multivariate exponential distribution. J. Amer. Statist. Assoc. **62**, 30–44 (1967)

25. Marshall, A.W., Olkin, I., Arnold, B.C.: Inequalities: theory of majorization and its applications, second edn. Springer Series in Statistics. Springer, New York (2011).
26. Mazo, G., Girard, S., Forbes, F.: A nonparametric class of one-factor copulas to balance flexibility and tractability. Tech. rep. (2013). HAL:hal-00979147
27. Mazo, G., Girard, S., Forbes, F.: Weighted least-squares inference based on dependence coefficients for multivariate copulas (2014). http://hal.archives-ouvertes.fr/hal-00979151
28. Muliere, P., Scarsini, M.: Characterization of a Marshall-Olkin type class of distributions. Ann. Inst. Statist. Math. **39**(2), 429–441 (1987)
29. Pinto, J., Kolev, N.: Extended Marshall–Olkin model and its dual version. In: U. Cherubini, F. Durante, S. Mulinacci (eds.) Marshall–Olkin Distributions – Advances in Theory and Practice
30. Salvadori, G., De Michele, C., Durante, F.: On the return period and design in a multivariate framework. Hydrol. Earth Syst. Sci. **15**, 3293–3305 (2011)
31. Salvadori, G., Durante, F., De Michele, C.: Multivariate return period calculation via survival functions. Water Resour. Res. **49**(4), 2308–2311 (2013)
32. Sklar, A.: Fonctions de répartition à *n* dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris **8**, 229–231 (1959)
33. Tankov, P.: Improved Fréchet bounds and model-free pricing of multi-asset options. J. Appl. Probab. **48**(2), 389–403 (2011)